Inventory and Monitoring Program

National Park Service
U.S. Department of the Interior

DATA MANAGEMENT PLAN: HEARTLAND I&M NETWORK AND PRAIRIE CLUSTER PROTOTYPE MONITORING PROGRAM

Gareth A. Rowell, Michael H. Williams,
and Michael D. DeBacker

Heartland I&M Network and Prairie
Cluster Prototype Monitoring Program
Wilson's Creek National Battlefield
Republic, Missouri

March 8, 2005

Executive Summary

To carry out its mission, the National Park Service (NPS) serves as a steward for the preservation of America's national parks and their resources. The NPS has initiated a service-wide, network-based inventory and monitoring (I&M) program to address the need for science-based natural resource information available to parks. The purpose of the program is to design and implement long-term ecological monitoring and provide information for park managers to evaluate the integrity of park ecosystems and better understand ecosystem processes.

The NPS is implementing park inventories and vital signs monitoring programs in approximately 270 natural resource parks. The NPS has organized these parks into 32 networks, linked by geography and shared natural resource characteristics. Parks in each network will share resources and professional expertise to implement a core program of inventory and monitoring that focuses on the key indicators of change or "vital signs" identified for the network's parks.

The Heartland I&M Network and Prairie Cluster Prototype Monitoring Program (HTLN) is composed of 15 parks in eight Midwestern states (Arkansas, Indiana, Iowa, Kansas, Minnesota, Missouri, Nebraska, Ohio) representing short/tallgrass prairie, Ozark highlands, and eastern deciduous forest eco-regions. Parks encompass a geographic area of over 720,000 square miles, include approximately 245,000 acres, and span roughly 2,700 feet of vertical relief.

This Data Management Plan identifies key data resources and processes to manage both inventory and monitoring data. A goal of this Plan is to describe the management of long-term monitoring data. Another goal of this Plan is to describe the resources and processes used to ensure high quality data. Specific objectives of the plan are:

- To describe the data management process;
- To describe roles and responsibilities of network staff and cooperators for managing data;
- To describe the current hardware and software environment in which we manage data;
- To summarize data and associated metadata collected and/or managed by the Heartland I&M Network and Prairie Cluster Prototype Monitoring Program;
- To schedule routine summary reports and trend analysis reports;
- To define future direction of data management activities in a work plan.

Network I&M staff and their cooperators make thousands of observations each year about plant and animal populations, communities and their environments. Taken together, these observations form a statistical representation of our sampling universe. In essence, the purpose of data management is to ensure that an accurate, complete record of those observations be maintained in perpetuity. Specific data management themes follow: data management infrastructure, data integration and exploration, and long-term data integrity and security. The focus of data management efforts within the Network is strategic within the context of service-wide data management activities. The service-wide Inventory and Monitoring Program has invested considerable resources in developing tools to archive and disseminate data (e.g. NPSpecies, Dataset Catalog, NatureBib). Further general guidance regarding data management procedures for handling and validating data is contained in the Data Management Plan: Prairie Cluster Prototype LTEM Program (DeBacker et al. 2002) and the Draft Data Management Protocol (Tessler & Gregson 1997). The Network data management system builds on

these resources while emphasizing and expanding the roles of geospatial database development and management during data collection and handling.

Assuring and maintaining data integrity is fundamental to the mission of an inventory and monitoring program and requires a considerable investment of staff time. The Data Manager, Project Manager, and Project Cooperator(s) share responsibility for data management within the Network. The Data Manager serves as GIS Specialist and Data Manager for the Network; the Project Manager supervises individual projects; the Project Cooperator carries out specific projects for the Network. Many of the Project Manager's duties involve close interaction with the Data Manager and Project Cooperator(s). In addition, the project manager will be the project Contracting Officers Technical Representative and/or designate a park Point-of-Contact to coordinate the project on behalf of the cooperating entity. Cooperating researchers have played an especially significant role in inventory projects and their associated data acquisition and data development. The cooperator Point-Of-Contact should be a permanent faculty member or agency employee who is responsible for the final delivery of data products.

In the interest of providing a more complete picture of our current project operations, we provide a list of hardware and software resources necessary to accomplish the goals and objectives as stated above. The automation equipment consists of approximately 10 Dell PC workstations and 1 mid-range dual CPU server. There has been an emphasis on Microsoft Access as software for database management. It is widely available, inexpensive, and interfaces with ESRI ArcGIS software to create geodatabases. Finally, we use SAS and NCSS for statistical analysis of monitoring data.

Data are direct measures of the population, community, or resource that is the focus of the protocol (for example, species inventory occurrence and abundance, rare species monitoring). Inventory data serves as the basis for future vital sign monitoring by identifying the fauna and flora. Monitoring data represent the change in vital signs over time. These vital signs follow sampling requirements defined in the monitoring protocol. For each Network, the project database is actually composed of many separate yet complimentary data sets. Each database should have event data and location data to reference observations in a specific time and place (i.e Natural Resource Database Template). Supplementary/ancillary environmental data assists in characterizing the specific environment or habitat that is associated with the primary vital sign. Metadata, developed as part of the data development process, also undergoes independent review to ensure compliance with NPS standards. Metadata is distributed at the NR-GIS clearinghouse with the I&M data sets, thus ensuring long-term value of the data. Additional information regarding GIS and metadata requirements, and metadata tools obtained at the NPS Midwest Region GIS website.

Almost all inventory and monitoring databases include a spatial component. The spatial data are usually represented as point data but may also be represented as line features (as in the case of transects) or polygons (for example, delineating habitat extents). ArcGIS personal geodatabases can be used as containers for both spatial and non-spatial data. The ArcGIS personal geodatabase allows shape files and coverages, representing GPS monitoring data, to be converted into feature classes. A feature class is a spatial data layer (point, line or polygon) stored in a relational database.

Data management procedures follow five key steps: acquisition, verification, validation, analysis and dissemination. In addition, storage, maintenance and security issues apply to all stages of the data flow. Data acquisition of Network I&M databases may include acquisitions through cooperative agreements with government agencies such as USGS-BRD or regional universities. Development of the Network natural resource knowledge base involved an initial phase of data mining of existing documents and reports. Data mining included bibliographic lists of reports about monitoring conducted within and around each park. Additional data mining captured park related monitoring, especially that conducted by neighboring agencies. These reports were scanned and optical character recognition of legacy documents was completed and then word-indexed. A database then summarized available information about monitoring activities within and around the Network parks.

Data acquisition may also be the direct result of field studies conducted by Network staff. Data acquisition may require data conversion from analog (usually paper) to digital or, on other occasions, be collected in digital form (GPS units and data loggers). Cooperators should obtain NPS park permits prior to initiating I&M studies, and both NPS staff and cooperators should obtain state and Federal fish and wildlife scientific permit letters prior to monitoring state or federally-listed species. All parties involved with any Network I&M project must be aware of any compliance issues regarding NPS Director's Orders.

Data entry is the initial set of operations in which raw data from paper field forms or field notebooks are transcribed into a computerized form (i.e., within a database). Data entry commences as soon as data collection is complete. Data entry forms and QA/QC features minimize error. Data entry forms reduce transcription errors through pick lists and value limits and provide controlled access to the database. Forms also control the sequence of data entry.

Data verification immediately follows data entry and involves checking the accuracy of computerized records against the original source—usually paper field records. While the goal of data entry is to achieve 100% correct entries, this is rarely accomplished. To minimize transcription errors, our policy is to verify all records to their original source by permanent staff. Further, 10% of records are reviewed a second time by the Project Manager and the results of that comparison reported with the data. The entire data set is re-verified if any errors are found. Once the computerized data are verified as accurately reflecting the original field data, the paper forms are scanned, archived and the electronic version is used for all subsequent data activities.

The process of reviewing computerized data for range and logic errors is the validation stage. Although data may be correctly transcribed from the original field forms, they may not be accurate or logical (eg. a stream pH of 25.0 or a temperature of 95°C is illogical). Certain components of data validation are built into data entry forms (e.g. range limits). Additional data validation can be accomplished during verification, if the operator is sufficiently knowledgeable about the data. The Project Manager will validate the data after verification is complete. Validation procedures seek to identify generic errors (e.g. missing, mismatched or duplicate records) as well as errors specific to particular projects. For example, validation of plant community data includes database query and comparison of data among years. The Project Manager must assure consistency between field forms and the database by noting how and why any changes were made to the data on the original field forms. In general, changes are made to field forms through marginal notes or attached explanations,

not erasures. Once validation is complete, the data set is turned over to the Data Manager for archiving and storage.

Secure data archiving is essential for protecting data files from corruption. Once a data set has passed the QA/QC procedures specified in the protocol, a formal entry is made in the I&M Dataset Catalog. Project Managers maintain the current data files for their project. These data files are stored on the program server. The server uses a RAID 5 configuration to ensure data integrity. Access to these files is controlled through the use of assigned user privileges. Typically, the Data Manager has universal access to data sets, while access for Project Managers is limited to their projects. Prior to storage on the server all files are scanned for viruses. Additional security is provided through tape backups. A complete server dump to tape is completed at least two times per week and three tape copies are stored locally and four off-site. Tape copies are rotated as needed.

The process of data analysis follows, but is not limited to, five basic steps: data determination, exploratory data analysis, the statistical test, a posteriori power analysis, and interpretation of the results. Data determination is comprised of several questions: Are data discrete (i.e. nominal or ordinal) or continuous (i.e. interval or ratio)? How were data collected? How many groups? And most importantly, what is the question being asked of the data? Proper use of data analysis will allow the ability to detect a change in the monitoring program if a change has occurred and ensure long-term viability of the monitoring program.

Efficient reporting is important to encourage the use of monitoring data in management decisions. To promote efficient reporting, data management efforts during the summary and analysis phase focus on automation of routine reports. Following the appropriate review process, dissemination of data, reports, and other items (photos, sound recordings, etc.) will be in accordance with current NPS standards. Species lists and voucher data will be input to NPSpecies and ANCS+ and subsequently updated (if required), reports documented in NatureBib, and other natural resource data and GIS data forwarded to the Biodiversity Data Store and Natural Resource GIS Data Store. Network geodatabases will be forwarded to the I&M protocol website for distribution following QA/QC and metadata review. In addition, completion and review of monitoring protocols and standard operating procedures (SOPs) will, most likely, be required prior to data distribution.

All data collected are public property and subject to requests under the Freedom of Information Act (FOIA). The NPS policy regarding FOIA is discussed in Director's Order 66 (which is in draft form at this time). Exemptions from FOIA are discussed in detail in the Director's Order. Of specific concern to data managers of NPS inventory and monitoring data are: cave locations and resources, federally listed endangered or threatened animals and plants, and state listed endangered or threatened animals and plants. Certain cultural and water resource data are also exempt from FOIA so data managers need to become familiar with Director's Order 66 prior to distributing information.

HTLN is just one of the 32 I&M networks in the National Park Service. Based on the data resources description in this document, the Network may be managing as many as 80 databases (including both inventory and monitoring data). The underlying goal of the program is to

provide park managers with the scientific data they need to understand and manage park resources. It will be a challenge to meet this goal given the total number of databases. The WASO office has been developing the Vital Signs Framework in order to address these issues and bring a unifying structure to the I&M program (Oakley *et al.*, 2003; Fancy 2004).

Acknowledgements

This document is a revision of DeBacker, M. D., B. Witcher, and J. R. Boetsch (2002). Jennifer Haack provided new graphics for this revision. We wish to extend our thanks to Albert J. Rowell for biometric review in Appendix 1. We are in debt to each of the participants of the 2004 I&M Data Management Plan Workshop. Special thanks to all of the inventory and monitoring staff and partners who are making the Vital Signs Monitoring Program a reality in our National Parks.

Table of Contents

1 Introduction 1
1.1 Inventory and monitoring program 1
1.2 Data management goals and objectives 2
2. Personnel Resources 4
2.1 Data Manager 4
2.2 Project Leader 6
2.3 Cooperators 7
3. Computer Resources 7
3.1 Network, Hardware and Peripherals 8
3.2 Software 8
3.3 Digital Data Formats 8
4. Data Resources 9
4.1 Inventory Data 9
4.2 Monitoring Data 10
4.3 Database Components 10
4.4 Using ArcGIS Geodatabases 17
5. Data Management Process 19
5.1 Data Flow Model 19
5.2 Data Acquisition 20
5.3 Data Verification 21
5.4 Data Validation 22
5.5 Data Organization 23
5.6 Data Maintenance and Storage 24
5.7 Data Analysis 25
5.7.1 Overview 25
5.7.2 Analysis for Annual Reports 25
5.7.3 Long-term Analysis of Monitoring Data 26
5.8 Data Management Necessary for Analysis 29
5.8.1 Overview 29
5.8.2 Identifying Project Objectives and Data Products 29
5.8.3 Data extraction 32
5.8.4 Automated Reporting and Data Summaries 32
5.8.5 Exports to statistical packages and other software 33
5.8.6 Data flow for subsequent analysis by third parties 33
5.8.7 Specific examples 35
5.9 Dissemination 37
5.10 Data Ownership and FOIA 37
5.10.1 National Park Service Policy on Data Ownership 38
5.10.2 Establishing Data Ownership Guidelines 38
5.11 Data Distribution 39
5.11.1 Data Classification: protected vs. public 39
5.11.2 Access Restrictions on Sensitive Data 40
5.11.3 Public Access to Network Inventory and Monitoring Data 41
5.11.4 Data Availability 42
5.11.5 Data Acquisition Policy 42
6. Database integration – Role of the Vital Signs Framework 43
7. Work Plan 45
Literature Cited 47
Appendix 1 49